DOCKET NO: 5363-001-23

TITLE OF THE INVENTION

NETWORK ASSET TRACKER FOR IDENTIFYING USERS OF NETWORKED COMPUTERS

BACKGROUND OF THE INVENTION

Field of the Invention

The present invention generally relates to computer networks, and more particularly to apparatus, systems, methods and computer program products that provide security within such computer networks.

Related Art

In today's technological climate it is typical for an enterprise (*i.e.,* a business concern, corporation, institution, organization, government agency or the like) to own and operate one or more computer networks (*e.g.,* local area networks (LANs) and the like). These computer networks may be spread out over several offices, floors and/or buildings. Within these computer networks are large amounts of sensitive, proprietary (and sometimes, confidential) data. Thus, it is understandable that such enterprises are concerned with the security of their computer networks.

Regardless of the implementation of login/password schemes, unauthorized users inevitably obtain access to computer networks. In fact, even those users to whom access of computer networks are authorized (*e.g.,* employees, independent contractors, sub-contractors and the like), may often use such networks in an unauthorized manner. Further, a great deal of unauthorized activity centers around electronic mail ("e-mail"). For example, an unauthorized user, or an authorized user

acting in an unauthorized manner, may send an enterprise's confidential data to unauthorized persons or unauthorized computer systems via the world-wide, public Internet using e-mail.

Given the above-described problem, what is needed is an apparatus, system, method and computer program product for identifying users of networked computers. Today, the problem is typically solved by first referring to any existing cable plant documentation (if available) or physically tracing the cable to a physical location. Then, security or IT personnel must arrive at the physical location in order to physically identifying the offending user. The needed apparatus, system, method and computer program product, however, should analyze network e-mail traffic and map Internet Protocol (IP) addresses to end users (*i.e.,* identify the user of a specific IP address within the network). The needed apparatus, system, method and computer program product would result in lowered response time for identifying, locating and disabling computers that pose a security threat.

SUMMARY OF THE INVENTION

The present invention meets the above identified needs by providing an apparatus, system, method and computer program product for identifying users of networked computers. That is, in an embodiment, the present invention provides a network asset tracking system that maps end users to workstation Internet Protocol (IP) addresses by passively analyzing (existing) network traffic. The network asset tracking system of the present invention also provides, in an embodiment, a reporting of end user-to-IP address mappings via a database-backed Web application.

In an embodiment, the network asset tracking system of the present invention includes two components -- a name discovery system "back end" and an administrative Web application "front end." The name discovery system ("NDS") is a "sniffer" apparatus (*i.e.,* hardware) connected to the primary switch of the enterprise's LAN. The NDS apparatus captures and analyzes network traffic. The Web application is provided for administrators of the computer network to manage and correlate the data captured by the NDS and cross-correlates such data with the enterprise's directory data to map IP addresses to end users.

An advantage of the present invention is that it allows users of computers that pose a security threat to be identified with lowered response time for locating and disabling the suspect computer.

Another advantage of the present invention is that it maps a computer user's identity to an organization's directory information (*e.g.,* building, room, phone, *etc.*), allowing the physical location of a computer to be determined (*i.e.,* identifying a specific building and/or room). Thus, security threats addressed by the present invention not only include those by unauthorized users, but also Trojan horse-type attacks where physically locating such attacks are critical.

Another advantage of the present invention is that it provides identification of computer users who are using a computer network's assets inappropriately and it can also identify computer users and their organization within a company for Information Technology (IT) infrastructure accounting purposes. This advantage becomes clearer when considering the accounting problem faced by large enterprises who share a large common network infrastructure, yet attempt to allocate the costs of network maintenance and support to separate divisions or departments.

Yet another advantage of the present invention is that it can identify errors in existing cable plant (network) documentation. By providing the physical location of a network connection, combined with the IP address on the switch port in the network closet, the present invention enables documenting the last "hop" and auditing of such existing network documentation.

Further features and advantages of the present invention as well as the structure and operation of various embodiments of the present invention are described in detail below with reference to the accompanying drawings.

BRIEF DESCRIPTION OF THE FIGURES

The features and advantages of the present invention will become more apparent from the detailed description set forth below when taken in conjunction with the drawings.

Figure 1 is a block diagram illustrating an enterprise's local area computer network in which the present invention may be implemented according to one embodiment.

Figures 2 and 3A-D are flowcharts illustrating network asset tracking processes according to alternate embodiments of the present invention.

Figure 4 is a block diagram of an exemplary computer system useful for implementing the present invention.

DETAILED DESCRIPTION

I. Overview

The present invention is directed to an apparatus, system, method and computer program product for identifying users of networked computers.

In an embodiment, the present invention is provided to an enterprise as a solution for mapping Internet Protocol (IP) addresses to an organization's personnel using directory data and the contents of network traffic. First, the enterprise's local area network (*e.g.*, Ethernet, FDDI or the like) traffic is captured and analyzed by installing a name discovery system apparatus (i.e., "NDS" hardware) on the primary switch of the enterprise's local area network (LAN). The captured data is cross-correlated with list data to map IP addresses to end users. Second, the network asset tracking solution of the present invention also provides access and manipulation of the collected network traffic data through a database-backed Web application for use by the enterprise's IT administrative personnel.

The apparatus, system, method and computer program of the present invention allow users of computers that pose a security threat to be identified with lowered response time for locating and disabling the suspect computer. Further, the present invention also allows an enterprise to perform accounting functions. For example, an enterprise may be interested in determining the network usage (*e.g.*, number of network connections) for a subset of computer users (*e.g.*, sub-contractors versus employees) for billing and other accounting purposes (*e.g.*, shared/allocated network infrastructure cost models employed by certain enterprises such as government agencies).

The present invention is now described in detail below in terms of the above examples. This is for convenience only and is not intended to limit the application of the present invention. In fact, after reading the following description, it will be apparent to one skilled in the relevant art(s) how to implement the following invention in alternative embodiments (*e.g.*, the analysis of different types of network traffic within different types of computer networks).

The terms "user," "entity," "personnel," "staff," "organization," "enterprise" and the plural form of these terms are used interchangeably throughout herein to refer to those who would access, use, be identified by and/or benefit from the tool that the present invention provides for identifying users of networked computers.

II. Apparatus and System Architecture

Referring to Figure 1, a network asset tracking ("NAT") system 100 according to an embodiment of the present invention is shown.

System 100 includes an enterprise's local area network (*e.g.*, Ethernet) backbone 102 which interconnects a plurality of end-user computers 104. In alternate embodiments, computers 104 are terminals, workstations (*e.g.*, Sun® SPARC™ or NT™ workstation running the Sun® Solaris™, Microsoft® Windows 2000™ or XP™, or IBM® AIX™ operating system) or personal computers (PC) (*e.g.*, an IBM™ or compatible PC running the Microsoft® Windows 95/98™ or Windows NT™ operating system, Macintosh® computer running the Mac® OS operating system, or the like). (For simplicity, Figure 1 shows computers 104a-n). In alternative embodiments, users may access LAN 102 using any processing device 104 including, but not limited to, a

desktop computer, laptop, palmtop, set-top box, personal digital assistant (PDA) and the like.

The backbone of LAN 102 is connected to a primary switch (*i.e.*, the LAN's primary Internet link) 106. Switch 106 is connected to a router 108 which in turn provides users of computers 104 with a connection to the public, global Internet 112.

In an embodiment, a name discovery system ("NDS") apparatus 110 is connected to primary switch 106. NDS 110 functions as a "sniffer" hardware (*i.e.*, a collection node) for capturing LAN 102 inbound and outbound traffic.

In one embodiment, NDS 110 is a one rack unit (1U) box with a power plug. In such an embodiment, NDS 110 has two 100Mbps network connections to primary switch 106. As shown in Figure 1, one link is a mirrored uplink, via one NDS 110 port to collect data from LAN 102. A second NDS 110 port is utilized for sending periodic data files and permitting regular access via a Web application. As will be appreciated by one skilled in the relevant art(s) after reading the description herein, in such an embodiment, NDS 110 requires two valid IP addresses. As will also be appreciated by those skilled in the relevant art(s) after reading the description herein, for larger networks, an NDS 110 can be installed at each core network uplink point (*i.e.*, primary switch) in an alternate embodiment.

In an embodiment, administrators of LAN 102 are given access to NDS 110 via a "front end" Web application which includes a login/password scheme. Such a front end is provided by Web server computer 114 having LAN 102 connectivity to NDS 110. As will be appreciated by one skilled in the relevant art(s), Web server 114 provides the "front-end" for NAT system 100. That is, server 114 contains a Web server process which sends out Web pages in response to Hypertext Transfer Protocol

(HTTP) or Hypertext Transfer Protocol (HTTPS) requests from remote browsers (*e.g.*, administrators of LAN 102). More specifically, it provides graphical user interface (GUI)"front-end" screens to such administrative users of NAT system 100 in the form of Web pages. These Web pages, when sent to the users' respective computers 104, result in GUI screens being displayed.

In an alternate embodiment, administrators of LAN 102 are also given remote access to NDS 110 via the Secure Shell (SSH) program on port 22 of the NDS 110.

As will also be appreciated by one skilled in the relevant art(s) after reading the description herein, in alternate embodiments, NDS 110 would contain, or have access to within NATS system 100, a central repository for storing all LAN 102 traffic data collected. Such a repository would also be accessible to the "front end" Web application to allow administrators of LAN 102 to collect statistics, view reports and the like.

More detailed descriptions of NAT system 100 components, as well their functionality, are provided below.

III. Operation

Referring to Figure 2, a flowchart illustrating the data flow of a network asset tracking process 200 according to an embodiment of the present invention is shown.

First, inbound and outbound e-mail traffic data 202 (*e.g.*, IP addresses and e-mail addresses) within LAN 102 are collected (*i.e.*, extracted) and stored by NDS 110. In an embodiment, the Tethereal ("dump and analyze network traffic") network protocol analyzer utility (developed as open source for Unix and Windows and available under the GNU General Public License) is used by NDS 110 to extract data

from LAN 102. In alternate embodiments, as will be appreciated by those skilled in the relevant art(s) after reading the description herein, other widely-available utilities (such as Snoop, Tcpdump or the like, or custom code logic) may be used by NDS 110 to extract data from LAN 102.

Next, Web server computer 114 (providing the above-mentioned database-backed Web application), having LAN 102 connectivity to NDS 110 would join the NDS 110 collected data and the enterprise's personnel directory information 206 in order to identify the users of computers 104 within LAN 102 (*i.e.,* map users to IP addresses). More specifically, server 114 provides GUI 208 "front-end" screens to such administrative users of NAT system 100 in the form of Web pages. These Web pages, when sent to the users' respective computers, result in GUI screens 208 being displayed.

In an embodiment, the enterprise's personnel directory information 206 is organized as an ITU-T X.500 or other formatted database containing data about the enterprise's personnel (*i.e.,* those authorized to use computers 104 within LAN 102). In an embodiment, such a database is a comma or tab delimited text file containing the exemplary fields listed in Table 1.

Enterprise Personnel Directory 206 Example Fields
First Name
Last Name
Middle Initial
Nick Names
Name Aliases
Building
Room
Permanent E-mail
Temporary E-mail
User Name
E-mail Address
Affiliation/Organization

Table 1

In an embodiment, NAT system 100 would generate, on a periodic time interval basis (*e.g.*, hourly, daily, weekly, *etc.*), an output data file containing all LAN 102 traffic data collected. In such an embodiment, the processing of data within NAT system 100 creates a text data file that is comma delimited for easy importing into other software application products (*e.g.*, Microsoft® Excel and the like). In alternate embodiments, the NAT system 100 output data file contains a subset or all of the exemplary fields listed in Table 2:

Example NAT Output File Fields
IP address
Hostname
First Name
Middle Initial
Last Name
E-mail Address
Location
Phone Number

Table 2

In an embodiment, the Web application GUI screens 208 provide the capability of sorting the tabular results on any returned field from Table 2. As will be appreciated by one skilled in the relevant art(s) after reading the description herein, the fields from Table 2 that can actually be presented in NAT system 100 output data files, and the resulting mapping of users to IP addresses, is dependent on the quality of the data found within the enterprise's personnel directory 206. As will also be appreciated by one skilled in the relevant art(s) after reading the description herein, Table 1 and Table 2 can be joined together using the E-mail Address field common to both tables.

It should be understood that Figure 2, which highlights the functionality and other advantages of NAT system 100, is presented for example purposes only. The architecture of the present invention is sufficiently flexible and configurable such that data collection and processing within NAT system 100 may take place in ways other than that shown in Figure 2 (*e.g.*, one or more data processing functions shown to take place on Web server 114 may take place on NDS 110 and *vice versa*).

VI. NDS Data Extraction

In an embodiment, NDS 110 is able to extract e-mail addresses and IP addresses from LAN 102 traffic data by analyzing port 25 of switch 106 for Simple Mail Transfer Protocol (SMTP) data, port 110 of switch 106 for Post Office Protocol, version 3 (POP3) data and port 143 of switch 106 for Internet Message Access Protocol, version 4 (IMAP) data.

Referring to Figure 3A, a flowchart illustrating the data flow of network asset tracking process 200 according to one embodiment of the present invention is shown. More specifically, in Figure 3A, computer 104 users are identified by NDS 110 from SMTP data traffic 202 exchanged between an enterprise's internal (SMTP) mail server 302 and external users 306 accessing outside (*e.g.*, public Internet) SMTP mail servers 308.

Most installations of SMTP servers do not implement data compression or encryption. The initial SMTP greeting will identify the domain from which the e-mail is originating. As seen in Figure 3A, extracted data 304 (*i.e.*, e-mail traffic data extracted by NDS 110) is analyzed by process 200. The command "MAIL FROM:" will identify the full e-mail address of the sender, and the command "RCPT TO:" will identify the full e-mail address of the recipient. Once NDS 110 extracts data from LAN 102, code logic stored therein is utilized to search for the following patterns to obtain user identifiers:

Command: MAIL

Request parameter: FROM:

or:

Command: RCPT

Request parameter: TO:

User identifiers will follow "FROM:" and "TO:" with the identifiers possibly contained with-in "<" and ">" characters. Words after the ":" and before a "<" will usually be some string of the user identifiers. ("FROM:" and "TO:" refer to sender and recipient, respectively.)

Referring to Figure 3B, a flowchart illustrating the data flow of network asset tracking process 200 according to one embodiment of the present invention is shown. More specifically, in Figure 3B, computer 104 users are identified by NDS 110 from POP3 traffic 202 exchanged between an enterprise's internal (POP) mail server 302 and external users accessing outside (*e.g.*, public Internet) mail servers.

The POP3 protocol does not use data encryption or compression. As seen in Figure 3B, extracted data 304 (*i.e.*, e-mail traffic data extracted by NDS 110) is analyzed by process 200. In POP3, a "USER" command is followed a space then the user identity (normally the username part of an e-mail address). Most implementations of the POP3 will usually have the "PASS" command follow the "USER" command. A "PASS" command will be followed by a space then the user's password in clear (*i.e.*, unencrypted text). A server response of "OK" will confirm the user's authenticity. Thus, in such an embodiment, a real-time analysis on the POP3 protocol is done using code logic to perform pattern matching for the following:

Request: USER

Request Arg:

"Request Arg:" will be followed by a username string that will identify a user's identity. With this information, the packet header will include source and destination

IP addresses to clearly identify the system the user is using. The inventor has found that, generally, less than 64 bytes of data is needed to capture the user's identifier.

Referring to Figure 3C, a flowchart illustrating the data flow of network asset tracking process 200 according to one embodiment of the present invention is shown. More specifically, in Figure 3C, computer 104 users are identified by NDS 110 from IMAP traffic 202 exchanged between an enterprise's internal (IMAP) mail server 302 and external users accessing outside (*e.g.,* public Internet) e-mail.

Like POP3, IMAP does not have data encryption or compression by default. As seen in Figure 3C, extracted data 304 (*i.e.,* e-mail traffic data extracted by NDS 110) is analyzed by process 200. Thus, a pattern match for the string "LOGIN" (case insensitive) will be used to identify a user's identity. After a "LOGIN" command has been issued to the server, a response of "OK LOGIN completed" or "FAIL" will confirm the user's identity. Obtaining a user's username for an IMAP system is similar to that of POP3 by examining for a pattern:

Request Tag: 000A

Request: LOGIN

Following the keyword "LOGIN" will be two arguments (username and password) wrapped in double quotes. Extracting only the necessary information, username, is done at this step. Similar to POP3, the inventor has found that less than 64 bytes of data is needed to be captured to obtain the user identifier. Depending on the client, the LOGIN command is normally within the first five IMAP packets sent.

Referring to Figure 3D, a flowchart illustrating the data flow of network asset tracking process 200 according to one embodiment of the present invention is shown. More specifically, in Figure 3D, computer 104 users are identified from Microsoft®

Exchange e-mail data traffic 202 exchanged between an enterprise's internal (Exchange) mail server 302 and external users 306 accessing outside (*e.g.*, public Internet) e-mail servers (not shown in Figure 3D).

Microsoft® Exchange Server 2000, and subsequent updates, encrypt traffic between Microsoft® Outlook clients (executing on the client computers 104) and the Exchange mail server 302. Thus, in an alternate embodiment of the present invention, a small script loaded on Exchange server 302 is utilized to obtain extracted data 304. That is, the script is executed at a pre-configured, regular interval, and leverages the Exchange Server 2000 Message Tracking Center (i.e., enabling the message tracking feature on server 302) and its associated tracking log files (e.g., yyyymmdd.txt) which reside on a server 302 share to extract IP and e-mail addresses of senders of e-mail within the network.

In an alternate embodiment, the Microsoft Exchange tracking log files can be remotely accessed using a script that leverages the filesystem object to open the log files and parse them to obtain IP and e-mail addresses of e-mail senders within the network.

In either of the two above-described embodiments, as seen in Figure 3D, extracted data 304 can then analyzed by process 200 as explained above. As will be appreciated by those skilled in the relevant art(s) after reading the description herein, the two above-described alternate embodiments leverage Exchange log files and thus allow NDS 110 to remain unutilized in such embodiments.

It should be understood that Figures 3A-D, which highlight the functionality and other advantages of NAT system 100, are presented for example purposes only. The architecture of the present invention is sufficiently flexible and configurable such

that data collection and processing within NAT system 100 may take place in ways other than that shown in Figures 3A-D.

V. Example Implementations

The present invention (system 100, process 200 or any part(s) or function(s) thereof) may be implemented using hardware, software or a combination thereof and may be implemented in one or more computer systems or other processing systems. In fact, in one embodiment, the invention is directed toward one or more computer systems capable of carrying out the functionality described herein. An example of a computer system 400 is shown in Figure 4. The computer system 400 includes one or more processors, such as processor 404. The processor 404 is connected to a communication infrastructure 406 (*e.g.*, a communications bus, cross-over bar, or network). Various software embodiments are described in terms of this exemplary computer system. After reading this description, it will become apparent to a person skilled in the relevant art(s) how to implement the invention using other computer systems and/or architectures.

Computer system 400 can include a display interface 402 that forwards graphics, text, and other data from the communication infrastructure 406 (or from a frame buffer not shown) for display on the display unit 430.

Computer system 400 also includes a main memory 408, preferably random access memory (RAM), and may also include a secondary memory 410. The secondary memory 410 may include, for example, a hard disk drive 412 and/or a removable storage drive 414, representing a floppy disk drive, a magnetic tape drive, an optical disk drive, *etc.* The removable storage drive 414 reads from and/or writes

to a removable storage unit 418 in a well known manner. Removable storage unit 418 represents a floppy disk, magnetic tape, optical disk, *etc.* which is read by and written to by removable storage drive 414. As will be appreciated, the removable storage unit 418 includes a computer usable storage medium having stored therein computer software and/or data.

In alternative embodiments, secondary memory 410 may include other similar devices for allowing computer programs or other instructions to be loaded into computer system 400. Such devices may include, for example, a removable storage unit 422 and an interface 420. Examples of such may include a program cartridge and cartridge interface (such as that found in video game devices), a removable memory chip (such as an erasable programmable read only memory (EPROM), or programmable read only memory (PROM)) and associated socket, and other removable storage units 422 and interfaces 420, which allow software and data to be transferred from the removable storage unit 422 to computer system 400.

Computer system 400 may also include a communications interface 424. Communications interface 424 allows software and data to be transferred between computer system 400 and external devices. Examples of communications interface 424 may include a modem, a network interface (such as an Ethernet card), a communications port, a Personal Computer Memory Card International Association (PCMCIA) slot and card, *etc.* Software and data transferred via communications interface 424 are in the form of signals 428 which may be electronic, electromagnetic, optical or other signals capable of being received by communications interface 424. These signals 428 are provided to communications interface 424 via a communications path (*e.g.*, channel) 426. This channel 426 carries signals 428 and

may be implemented using wire or cable, fiber optics, a telephone line, a cellular link, an radio frequency (RF) link and other communications channels.

In this document, the terms "computer program medium" and "computer usable medium" are used to generally refer to media such as removable storage drive 414, a hard disk installed in hard disk drive 412, and signals 428. These computer program products provide software to computer system 400. The invention is directed to such computer program products.

Computer programs (also referred to as computer control logic) are stored in main memory 408 and/or secondary memory 410. Computer programs may also be received via communications interface 424. Such computer programs, when executed, enable the computer system 400 to perform the features of the present invention, as discussed herein. In particular, the computer programs, when executed, enable the processor 404 to perform the features of the present invention. Accordingly, such computer programs represent controllers of the computer system 400.

In an embodiment where the invention is implemented using software, the software may be stored in a computer program product and loaded into computer system 400 using removable storage drive 414, hard drive 412 or communications interface 424. The control logic (software), when executed by the processor 404, causes the processor 404 to perform the functions of the invention as described herein.

In another embodiment, the invention is implemented primarily in hardware using, for example, hardware components such as application specific integrated circuits (ASICs). Implementation of the hardware state machine so as to perform the functions described herein will be apparent to persons skilled in the relevant art(s).

In yet another embodiment, the invention is implemented using a combination of both hardware and software.

VI. Conclusion

While various embodiments of the present invention have been described above, it should be understood that they have been presented by way of example, and not limitation. It will be apparent to persons skilled in the relevant art(s) that various changes in form and detail can be made therein without departing from the spirit and scope of the present invention. Thus, the present invention should not be limited by any of the above-described exemplary embodiments, but should be defined only in accordance with the following claims and their equivalents.